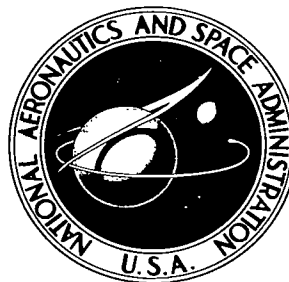


NASA TECHNICAL NOTE



NASA TN D-3481

NASA TN D-3481

LOAN COPY: RETURN
AFWL (WLIL-2)
KIRTLAND AFB, NM



THE BIASING EFFECT OF RANDOM BIT ERRORS ON BINARY TELEMETRY DATA

by Morton Pasternack

*Goddard Space Flight Center
Greenbelt, Md.*



THE BIASING EFFECT OF RANDOM BIT ERRORS
ON BINARY TELEMETRY DATA

By Morton Pasternack

Goddard Space Flight Center
Greenbelt, Md.

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

For sale by the Clearinghouse for Federal Scientific and Technical Information
Springfield, Virginia 22151 - Price \$1.00

ABSTRACT

Random bit errors that occur in pulse code modulated (PCM) data during the various phases of the telemetry-data-processing system will not only add a component of mean square error to the experimental data, but will also bias them. This paper derives a formula for the bias and describes some of its characteristics.

CONTENTS

Abstract	ii
INTRODUCTION	1
BIASED VS. UNBIASED ESTIMATES	2
MEAN SQUARE ERROR VS. BIAS ERROR	2
DERIVATION OF THE BIAS FORMULA	2
Assumptions	2
Symbol Definitions	2
Derivation	3
SOME CHARACTERISTICS OF THE BIAS	5
TIME PLOT DISTORTIONS	7
CONCLUSION	8
References	8

THE BIASING EFFECT OF RANDOM BIT ERRORS ON BINARY TELEMETRY DATA

by

Morton Pasternack

Goddard Space Flight Center

INTRODUCTION

PCM telemetry data, after digitization at a data processing facility, will contain errors which were introduced at various phases of the telemetry-data-processing system. We may think of the error sources as being divided into two parts: (a) prior to transmission, and (b) during and after transmission. The errors occurring prior to transmission result from imprecise experimental measuring devices, sampling error, quantizing error, encoding errors, etc. The errors occurring during and after transmission will occur in the form of bit errors and result from transmission signal errors, and detection and decision process errors. It is the latter type error (errors occurring during and after transmission) that is the subject of this paper. The orientation of this paper is mainly towards the situation where data are being transmitted from an experiment aboard a PCM satellite, and the statistic of interest is the mean value of the data points* transmitted during a time interval of duration ΔT . It should be noted, however, that the findings of this paper are equally valid for the case where the statistic of interest comprises the individual data points themselves, rather than an average of data points.

Bit errors that occur during the transmission—detection—decision process will add a component of mean square error to the mean square error which occurred in the data prior to transmission. It is obvious that if the probability of occurrence of a bit error is correlated with the transmitted value of that bit (i.e., the probability of a transmitted "one" being received in error is not equal to the probability of a transmitted "zero" being received in error), then a biasing error as well as a component of mean square error will be introduced into the data. What is not obvious is that "randomly distributed"[†] bit errors will also result in a biasing type error, as well as in a mean square error.

*A data point is defined for the purpose of this paper as a single measurement of an experimental parameter. The value of a data point can be any value from 0 to $2^m - 1$, where m = the number of bits in a data channel.

[†]The term "randomly distributed" is used to describe the distribution of bit errors, for lack of a better short description. Each time it appears in this paper, its intended meaning is that the probability of any given received bit being in error is equal to the probability of any other received bit being in error (irrespective of the transmitted value of the given bit, the bit position within the data channel of the given bit, the occurrence or non-occurrence of previous bit errors, or any other attribute of the given bit).

BIASED VS. UNBIASED ESTIMATES

Bias is defined as the deviation of the *expected value* of an estimating procedure from the *true value*. The expected value in turn is the average estimate that would ultimately be approached if the estimating procedure were repeated indefinitely.

We can look upon $\bar{\alpha}$ (the mean transmitted value) as the true value, and the transmission-detection-decision process as the estimating procedure that produces the estimate $\bar{\beta}$ (the mean received value). $\bar{\beta}$ is then by definition an unbiased estimate of $\bar{\alpha}$, if the expected value of $\bar{\beta}$ is equal to $\bar{\alpha}$. Conversely, $\bar{\beta}$ is by definition a biased estimate of $\bar{\alpha}$, if the expected value of $\bar{\beta}$ does not equal $\bar{\alpha}$.

MEAN SQUARE ERROR VS. BIAS ERROR

Unlike the mean square error, the bias error is independent of n , the number of data point values (DPV's) used to compute $\bar{\beta}$. Whereas the mean square error of $\bar{\beta}$ can be reduced by increasing n , the bias error remains constant for all values of n . The mean square error results in an imprecise estimate and the bias type error results in a distorted estimate. If a time plot were made of the $\bar{\beta}$ values, curve fitting techniques could, for the most part, eliminate the effects of the mean square error, but they could not eliminate the effects of the bias.

Formulas for computing the *mean square error* resulting from randomly distributed bit errors have been derived previously by S. Karp (Reference 1), by M. B. Rudin and J. Siegel (Reference 2), and by others. It is the purpose of this paper to derive a formula for the *biasing type error* resulting from randomly distributed bit errors, and to describe some of the characteristics of this bias.

DERIVATION OF THE BIAS FORMULA

Assumptions

Two assumptions are made:

1. For each transmitted data point value (α_i), there is a corresponding received data point value (β_i).
2. P_e is constant for each of the $m \times n$ bits that comprise $\bar{\alpha}$.

Symbol Definitions

α_i = the i^{th} transmitted data point value ($0 \leq \alpha_i \leq 2^m - 1$)

β_i = the i^{th} received data point value ($0 \leq \beta_i \leq 2^m - 1$)

$$\bar{\alpha} = 1/n \sum_{i=1}^n \alpha_i$$

$$\bar{\beta} = 1/n \sum_{i=1}^n \beta_i$$

$E(\beta_i)$ = the expected value of β_i

$E(\bar{\beta})$ = the expected value of $\bar{\beta}$

n = the number of data point values transmitted

m = the number of bits in a data channel

P_e = the bit error probability (this is equal to the probability that the transmitted value of any given bit, zero or one, will differ from the received value).

j = the bit position within a data channel ($j = 0$ represents the low order bit position; $j = m - 1$ represents the high order bit position)

A_{ij} = the value (zero or one) of the j^{th} bit position in the binary representation of α_i

B_{ij} = the value (zero or one) of the j^{th} bit position in the binary representation of β_i

$P(B_{ij} = 1)$ = the probability that the value of B_{ij} is one

$P(B_{ij} = 0) = 1 - P(B_{ij} = 1)$ = the probability that the value of B_{ij} is zero.

Derivation

By definition,

$$\alpha_i = \sum_{j=0}^{m-1} A_{ij} 2^j ,$$

$$\beta_i = \sum_{j=0}^{m-1} B_{ij} 2^j ,$$

and

$$\begin{aligned}
 E(\beta_i) &= \sum_{j=0}^{m-1} E(B_{ij}) 2^j \\
 &= \sum_{j=0}^{m-1} [P(B_{ij} = 1)(1) + P(B_{ij} = 0)(0)] 2^j \\
 &= \sum_{j=0}^{m-1} 2^j P(B_{ij} = 1) .
 \end{aligned}$$

Now, $P(B_{ij} = 1)$ is equal to $1 - P_e$ when the value of the corresponding A_{ij} is one, and is equal to P_e when the value of the corresponding A_{ij} is zero. This can be expressed as

$$P(B_{ij} = 1) = (1 - P_e)(A_{ij}) + P_e(1 - A_{ij}) .$$

Therefore,

$$\begin{aligned}
 E \beta_i &= \sum_{j=0}^{m-1} 2^j [(1 - P_e)(A_{ij}) + P_e(1 - A_{ij})] \\
 &= \sum_{j=0}^{m-1} 2^j [P_e + (1 - 2P_e) A_{ij}] \\
 &= P_e \sum_{j=0}^{m-1} 2^j + (1 - 2P_e) \sum_{j=0}^{m-1} A_{ij} 2^j \\
 &= P_e (2^m - 1) + (1 - 2P_e) \alpha_i .
 \end{aligned}$$

Hence,

$$\begin{aligned}
 E(\bar{\beta}) &= E\left(\frac{1}{n} \sum_{i=1}^n \beta_i\right) = \frac{1}{n} \sum_{i=1}^n E(\beta_i) = \frac{1}{n} \sum_{i=1}^n [P_e (2^m - 1) + (1 - 2P_e) \alpha_i] \\
 &= P_e (2^m - 1) + (1 - 2P_e) \left(\frac{1}{n}\right) \sum_{i=1}^n \alpha_i = P_e (2^m - 1) + (1 - 2P_e) \bar{\alpha} .
 \end{aligned}$$

Thus,

$$\begin{aligned}
 \text{bias} &= E(\bar{\beta}) - \bar{\alpha} = P_e (2^m - 1) + (1 - 2P_e) \bar{\alpha} - \bar{\alpha} \\
 &= P_e (2^m - 1) - 2P_e \bar{\alpha} \\
 &= 2P_e \left(\frac{2^m - 1}{2} - \bar{\alpha} \right).
 \end{aligned}$$

SOME CHARACTERISTICS OF THE BIAS

From the formula

$$E(\bar{\beta}) - \bar{\alpha} = 2P_e \left(\frac{2^m - 1}{2} - \bar{\alpha} \right),$$

we can see the following:

- a. The absolute value of the bias is a function of the absolute difference between $(2^m - 1)/2$ and $\bar{\alpha}$, where $2^m - 1$ is the maximum DPV of an m bit data channel. When $\bar{\alpha} = (2^m - 1)/2$, the bias is equal to zero. When $\bar{\alpha} = 0$ or $2^m - 1$, the absolute value of the bias is a maximum.
- b. The bias is directly proportional to P_e .

A feel for the magnitude of the bias and the manner in which it varies can be derived by substituting numbers for m and P_e , and varying the value of $\bar{\alpha}$. The results are presented in Table 1. The values chosen for m and P_e are 9 and 0.01, respectively.

Referring to the second entry in Table 1, note that when $m = 9$, $P_e = 0.01$, and $\bar{\alpha} = 51.1$, then $E(\bar{\beta}) = 55.188$, and the bias = 4.088. This means that if an infinite number of DPV's whose average was 51.1 were transmitted by means of a 9-bit data channel, with a bit error probability of 0.01, then the average of the DPV's received would be 55.188. The bias would be $55.188 - 51.100 = 4.088$.

It was noted earlier that the bias is directly proportional to P_e . Therefore, using the example just given, when $\bar{\alpha} = 51.1$ and $m = 9$,

$$\text{the bias} = 40.88 \text{ when } P_e = 0.1,$$

$$\text{the bias} = 4.088 \text{ when } P_e = 0.01, \text{ and}$$

$$\text{the bias} = 0.4088 \text{ when } P_e = 0.001.$$

Figure 1 graphically portrays the relationship between $\bar{\alpha}$, $E(\bar{\beta})$, and $E(\bar{\beta}) - \bar{\alpha}$. An extremely high value for P_e was utilized ($P_e = 0.1$), in order to make the relationship on the chart legible. The vertical distance between the dashed line, $E(\bar{\beta})$, and the solid line, $\bar{\alpha}$, at any $\bar{\alpha}$ value, represents the bias for that $\bar{\alpha}$ value. For example, if $\bar{\alpha} = 102.2$, then $E(\bar{\beta}) = 132.86$ and the bias = 30.66.

Table 1

Relationship of Mean Transmitted DPV ($\bar{\alpha}$),
Expected Value of Mean Received DPV ($E(\bar{\beta})$),
and Bias ($E(\bar{\beta}) - \bar{\alpha}$), when $m = 9$, and $P_e = 0.01$.

$\bar{\alpha}$	$E(\bar{\beta})$	Bias = $E(\bar{\beta}) - \bar{\alpha}$
$0(2^m - 1) = 0$	5.110	5.110
$0.1(2^m - 1) = 51.1$	55.188	4.088
$0.2(2^m - 1) = 102.2$	105.266	3.066
$0.3(2^m - 1) = 153.3$	155.344	2.044
$0.4(2^m - 1) = 204.4$	205.422	1.022
$0.5(2^m - 1) = 255.5$	255.500	0
$0.6(2^m - 1) = 306.6$	305.578	-1.022
$0.7(2^m - 1) = 357.7$	355.656	-2.044
$0.8(2^m - 1) = 408.8$	405.734	-3.066
$0.9(2^m - 1) = 459.9$	455.812	-4.088
$1.0(2^m - 1) = 511.0$	505.890	-5.110

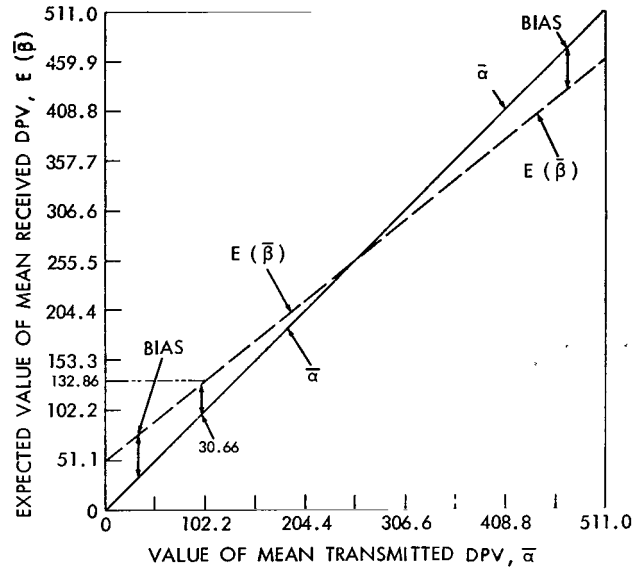


Figure 1—Graphic presentation of relationship between $\bar{\alpha}$, $E(\bar{\beta})$, and $E(\bar{\beta}) - \bar{\alpha}$ ($m = 9$, $P_e = 0.1$).

Table 2

Relationship of Mean Transmitted DPV ($\bar{\alpha}$), Bias ($E(\bar{\beta}) - \bar{\alpha}$), and Bias as a Percentage of $\bar{\alpha}$ [$100 (E(\bar{\beta}) - \bar{\alpha}) / \bar{\alpha}$] when $m = 9$ and $P_e = 0.01$.

$\bar{\alpha}$	Bias = $E(\bar{\beta}) - \bar{\alpha}$	$100 (E(\bar{\beta}) - \bar{\alpha}) / \bar{\alpha}$ (percent)
$0(2^m - 1) = 0$	5.1100	Not Defined
$0.005(2^m - 1) = 2.555$	5.0589	198
$0.01 (2^m - 1) = 5.11$	5.0078	98
$0.02 (2^m - 1) = 10.22$	4.9056	48
$0.05 (2^m - 1) = 25.55$	4.5990	18
$0.10 (2^m - 1) = 51.10$	4.0880	8
$0.20 (2^m - 1) = 102.20$	3.0660	3
$0.30 (2^m - 1) = 153.30$	2.0440	1.33
$0.40 (2^m - 1) = 204.40$	1.0220	0.50
$0.50 (2^m - 1) = 255.50$	0	0
$0.60 (2^m - 1) = 306.60$	-1.0220	-0.33
$0.70 (2^m - 1) = 357.70$	-2.0440	-0.57
$0.80 (2^m - 1) = 408.80$	-3.0660	-0.75
$0.90 (2^m - 1) = 459.90$	-4.0880	-0.89
$1.00 (2^m - 1) = 511.00$	-5.110	-1.00

Another important relationship, probably more important than the absolute bias, is the bias as a percentage of \bar{a} . Table 2 shows the relationship of this percentage to the value \bar{a} , for $m = 9$ and $P_e = 0.01$.

The figures in Table 2 give a good indication of the magnitude of the bias problem when P_e is high ($P_e = 0.01$). For example, if we have a 9-bit data channel, a 0.01-bit error probability, and a 10.22 value of \bar{a} , then the estimate of \bar{a} , based on the received values, would be biased by 48 percent. It should also be noted that the bias, as a percentage of \bar{a} , is directly proportional to P_e . Therefore, using the entry for $\bar{a} = 10.22$ in Table 2 as an example, when $\bar{a} = 10.22$ and $m = 9$,

the bias as a percentage of $\bar{a} = 480$ percent when $P_e = 0.1$,

the bias as a percentage of $\bar{a} = 48$ percent when $P_e = 0.01$,

the bias as a percentage of $\bar{a} = 4.8$ percent when $P_e = 0.001$, and

the bias as a percentage of $\bar{a} = 0.48$ percent when $P_e = 0.0001$.

TIME PLOT DISTORTIONS

The previous section of this paper dealt with the effect of randomly distributed bit errors on the relationship between a mean transmitted DPV and the corresponding mean received DPV. In this section, note is taken of randomly distributed bit errors and their distorting effect on time plots. Figures 2 and 3 graphically illustrate examples of this distortion. The charts are based on the following assumptions:

1. The transmitted DPV's for a given experiment, plotted on a time scale, take the form of a sine wave.
2. A time plot is made from the received DPV's. Each point on this time plot is made up of an average of n received DPV's during an increment of time equal to Δt .
3. Δt approaches 0 as n approaches ∞ .

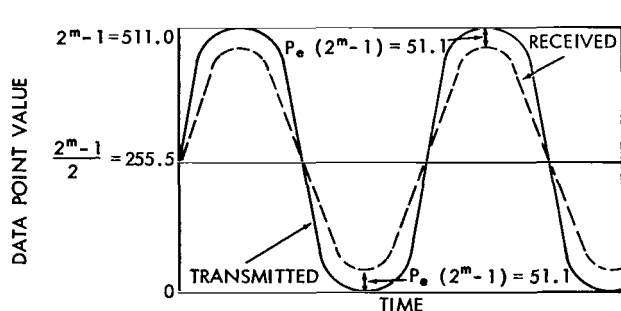


Figure 2—Transmitted sine wave (with amplitude = $(2^m - 1)/2$) vs. the received curve ($m = 9$, $P_e = 0.1$).

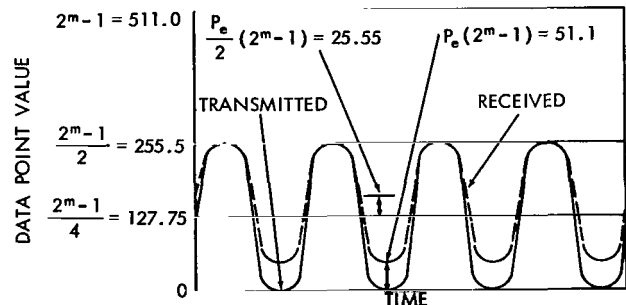


Figure 3—Transmitted sine wave (with amplitude = $(2^m - 1)/4$) vs. the received curve ($m = 9$, $P_e = 0.1$).

4. The received DPV's contain randomly distributed bit errors.
5. $m = 9$ and $P_e = 0.1$. (A high P_e was chosen in these illustrations in order to make the relationship on the charts legible.)

Figures 2 and 3 show the relationship of the transmitted curve versus the received curve. In both figures, the minimum value of the transmitted curve is zero. Figure 2 shows the relationship when the amplitude of the transmitted sine wave is equal to $(2^m - 1)/2$; Figure 3 shows the relationship when the amplitude of the transmitted sine wave is equal to $(2^m - 1)/4$.

In Figure 2, the transmitted curve fluctuates about $(2^m - 1)/2$, the mid-point of an m -bit data channel. When the transmitted curve is above $(2^m - 1)/2$, the received curve is below the transmitted curve. When the transmitted curve is below $(2^m - 1)/2$, the received curve is above the transmitted curve. In Figure 3, the transmitted curve is never above $(2^m - 1)/2$. The received curve is, therefore, never below the transmitted curve. In both Figure 2 and Figure 3, the maximum absolute bias occurs when the transmitted curve value is zero. At that point, the bias is equal to $P_e (2^m - 1)$. In both figures the value of the absolute bias decreases as the curve approaches $(2^m - 1)/2$, at which point the bias is zero.

CONCLUSION

Intuition might lead one to conclude that if bit errors were randomly distributed, the averaging of a large number of received DPV's would tend to cause the effects of the bit errors in one direction to cancel those in the opposite direction. It would follow that the mean received DPV could serve as an estimate of the mean transmitted DPV, and that this estimate would be unbiased. Unfortunately this would not be the case. The estimate would be a biased estimate and the amount of the bias would be equal to

$$2P_e \left(\frac{2^m - 1}{2} - \bar{\alpha} \right).$$

Curves developed from received DPV's which are biased, will be distortions of the curves described by the transmitted DPV's. The amount of distortion of a given segment of a curve will be dependent upon the value of P_e and of the difference between the value of the points on the curve and the value $(2^m - 1)/2$. The closer a segment of the transmitted curve is to the value $(2^m - 1)/2$, the less the distortion.

(Manuscript received January 7, 1966)

REFERENCES

1. Karp, S., *IEEE Trans. On Space Electronics and Telemetry* (Correspondence), Vol. SET-10, September 1964, p. 124.
2. Rudin, M. B., and Siegel, J., "Telemetry System Study, Final Rep't.," *Aeronautic Publication No. U-743*, Vol. 1, Appendix 2, December 1959.

"The aeronautical and space activities of the United States shall be conducted so as to contribute . . . to the expansion of human knowledge of phenomena in the atmosphere and space. The Administration shall provide for the widest practicable and appropriate dissemination of information concerning its activities and the results thereof."

—NATIONAL AERONAUTICS AND SPACE ACT OF 1958

NASA SCIENTIFIC AND TECHNICAL PUBLICATIONS

TECHNICAL REPORTS: Scientific and technical information considered important, complete, and a lasting contribution to existing knowledge.

TECHNICAL NOTES: Information less broad in scope but nevertheless of importance as a contribution to existing knowledge.

TECHNICAL MEMORANDUMS: Information receiving limited distribution because of preliminary data, security classification, or other reasons.

CONTRACTOR REPORTS: Technical information generated in connection with a NASA contract or grant and released under NASA auspices.

TECHNICAL TRANSLATIONS: Information published in a foreign language considered to merit NASA distribution in English.

TECHNICAL REPRINTS: Information derived from NASA activities and initially published in the form of journal articles.

SPECIAL PUBLICATIONS: Information derived from or of value to NASA activities but not necessarily reporting the results of individual NASA-programmed scientific efforts. Publications include conference proceedings, monographs, data compilations, handbooks, sourcebooks, and special bibliographies.

Details on the availability of these publications may be obtained from:

SCIENTIFIC AND TECHNICAL INFORMATION DIVISION
NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
Washington, D.C. 20546